

LATS Format and Convention Options: Pros and Cons

The two format options in LATS, netCDF and GRIB, have distinctly different advantages. AMIP II users are free to choose the LATS format which is most appropriate for their needs but are required to use either the LATS_COARDS or LATS_GRADS_GRIB conventions listed below. The LATS *Convention* defines how the data and the metadata (data on the data) are structured as well as the actual file format. It is the standards on metadata and structure that makes LATS data accessible to a data interface. Some factors to consider when choosing LATS format/convention options are summarized below. Note that the data format is implied by the convention.

Convention options

LATS_PCMDI (netCDF): Permits the use of calendars other than the standard Gregorian calendar. Other than the calendar options, this option is at present consistent with the COARDS conventions. Future releases may provide other extensions beyond the COARDS convention. Files created with this convention may be used with all netCDF utilities, except those that require the COARDS convention (e.g., GrADS). As this is an evolving convention, its use in AMIP II is not supported.

LATS_COARDS (netCDF): Compliant with the COARDS metadata standard. If this convention is chosen, the calendar must be specified to be Gregorian (LATS_STANDARD calendar). Internally, monthly data are represented in units of "days since (basetime)". Climatology data are assumed if year = 2. Fixed fields (e.g., topography) is assigned the year 1.

LATS_GRADS_GRIB: WMO GRIB format (GRIB), plus a GrADS control file and a GRIB map file. Variables must be defined on the same horizontal grid. Supports, all calendar options except Julian and 360-day years.

LATS_GRIB_ONLY: WMO GRIB only. The delta (see LATS man page) may be zero. This convention is the most general, but may not be readily interfaced to VCS and GrADS. Appropriate for straight data transmission with no need for direct data interface.

Data format pros and cons:

netCDF – advantages: There is a wide range of tools available for netCDF data, and it is used extensively in the atmospheric science research community, particularly among universities. The netCDF data model is general and flexible. All quality controlled AMIP data released to diagnostic subprojects will be in netCDF.

netCDF – disadvantages: Requires installation and use of the netCDF libraries. For some platforms, netCDF version 2.4 or greater must be used for performance reasons. At present, LATS does not support the COARDS data compression (offset plus scaling to 16-bit integers) option, but will do so in a future release. Data writing and reading can be much slower than GRIB, particularly for large grids (e.g., 360x180 points).

GRIB – advantages: GRIB is a WMO standard and is widely used at most numerical weather prediction centers (e.g., ECMWF and NCEP). GRIB features flexible compression options on a variable-by-variable basis and the binary format is generally simple. For AMIP II high frequency (6-h) data, GRIB files will be on average 2.5 times smaller than netCDF format, i.e., a space savings of ~60%.

GRIB – disadvantages: GRIB is primarily a physical format with metadata defined for each 2–D grid (i.e., a "GRIB message") only. Thus, GRIB data cannot be organized internally as higher dimensional data sets. The higher dimensional structure achieved in the LATS_GRADS_GRIB convention is though a mapping function not part of the format. For example, **all variables in a file must share the same vertical dimension, and multilevel variables must share the same vertical dimension.** Beyond the GrADS interface, there are no standard programming interfaces and a variety of GRIB 'flavors' (LATS uses WMO GRIB Edition 1 and the NCEP tradition) have evolved complicating interface building.

Design Constraints and Resulting Limitations

The LATS software library was designed under two overarching constraints — simplicity and two–format output. Supporting output of both netCDF and GRIB, from the same interface, required rectification to a common denominator. netCDF is a self describing, data–set orientated (e.g., 4–D variables) format whereas GRIB is only designed to output 2–D lon/lat grids. The GrADS interface allows GRIB data to be treated as if it were a data–set in the netCDF sense (i.e, the netCDF 4–D "data model"). Thus, to support the GrADS interface to GRIB and for consistency with netCDF, ***LATS requires all variables to be declared (including all their dimensions and attributes), before any numerical data are written.*** While netCDF does in fact allow an application to define and write in any order, additional functions would have to be added to the LATS API making the system more complex. Further, a relaxation of the define once, then write constraint would allow for a greater range of data structures (e.g, a set of time–series) and thus greater complexity in the processing at PCMDI. However, we expect to ease the define once, write many restriction in future versions.

Although four conventions may seem complex, it is symptomatic of the wide variation in existing data sets as conventions are often specific to an individual. The purpose of a convention is to improve the efficiency of a group and the lack of conventions for AMIP I was very costly to PCMDI as we had to bring the disparate data sets (all 30 model data sets were different) to some common standard. Relative to AMIP I, four conventions will greatly improve the efficiency of PCMDI and our ability to support analysis of the AMIP II results.

While the valid year of a climatology field may have no mean, it is generally required by data interfaces. LATS internally sets a time convention for climatology and fixed fields. Again, there are no obvious standards for describing the valid time of climatology or fields such topography. Thus, we have arbitrarily set the valid year of a data written under the climatology calendars to the year 2 AD and fixed fields to 1 AD.

CDAT, EZGET (a Fortran interface) and GrADS can read **both** GRADS_GRIB and netCDF (COARDS) LATS–generated data. FERRET, and many other applications can be used to access the netCDF (COARDS) LATS–generated data. EZGET users creating applications that link to both EZGET and LATS may experience problems because of compiler–dependent resolution of externals (functions). The issue is compatibility between cdunif and lats routines which is now being corrected at PCMDI.